Report to:
Social Services Appropriations

Department of Health
Master Person Index
(DOHMPI)

Prepared by the:
Utah Department of Health
Center for Health Data and Informatics
Health Informatics Office

REPORT PURPOSE

The purpose of this report is to respond to this legislative request regarding the Department of Health Master Person Index (DOHMPI):

The Legislature intends that the Department of Health report on the efficiencies, impacts, process changes, and accompanying cost impacts achieved by the Master Person Index project by January 2022 to the Social Services Appropriations Subcommittee.

SUMMARY AND KEY POINTS

DOHMPI provides efficiencies that allow the state to complete tasks more quickly or to complete tasks it may not otherwise have resources to perform by automating processes. DOHMPI does not replace program specific data systems, but serves as a data sharing hub connecting data systems so they can share information more seamlessly. These tasks include: person matching, so we know if the same person exists in one or more data systems; automated extract, transform, load (ETL) operations; messaging workflows that provide up-to-date person matching; and information sharing among data systems and for Person-Linked Data Warehouses.

Sixteen Department of Health (UDOH), Department of Workforce Services (DWS), Department of Professional Licensing (DOPL) and external data systems' identities are currently integrated with DOHMPI with more planned.

The Early Childhood Integrated Data Systems (ECIDS) is leveraging DOHMPI to populate its Person-Linked Data Warehouse to better serve Utah's early childhood population.

The DOHMPI data sharing hub infrastructure has already been built so additional UDOH and Department of Human Services (DHS) programs can be integrated with DOHMPI at an incremental cost. DOHMPI's role is being considered in the construction of the new Department of Health and Human Services (DHHS) data warehouse and for information sharing among DHHS programs.

To preserve privacy, systems are only linked and data is only shared as allowed by statute following minimum necessary standards.

OVERVIEW OF DOHMPI

Data Sharing Hub

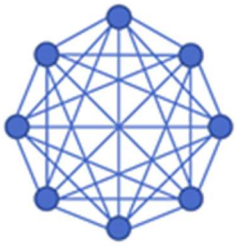
DOHMPI is like an airport hub (See figure 1 - Hub) for data integration and sharing for the Utah Department of Health (UDOH). This architecture helps reduce the complexity associated with point-to-point (See figure 2 - Point-to-Point) data connections and integrations. Fewer point-to-point routes results in lower development and maintenance costs. DOHMPI can be used for both operational and research use cases.

Figure 1 - Hub

Hub

linking 8 points, requires 8 routes

Figure 2 - Point-to-Point



Point-to-Point

linking 8 points, requires 28 routes

Each UDOH program has their own program system specific data silo. DOHMPI is an information sharing hub for UDOH programs and external UDOH Data Partners (e.g. Department of Workforce Services (DWS), Division of Occupational and Professional Licensing (DOPL) and private sector organizations), where sharing is allowed by law and necessary for specific business practices.

The DOHMPI database contains demographic data from contributing systems that is used to match people across systems. System specific data (e.g. immunizations, birth details) remains in the source system(s), but can be shared among the connected systems using DOHMPI.

As such DOHMPI is a system that will need ongoing funding to enhance and maintain the system to support the operations of the department. The system may also provide the basis for identity resolution in the new Department of Health and Human Services data warehouse.

A list of some of the DOHMPI functionality includes:

- **Real-time Identity Resolution and Record Linkage:** Using sophisticated matching algorithms, DOHMPI links people across independent systems. This means DOHMPI is able to identify if the same person exists in System A as in System B, System C, etc. This is helpful in understanding how many unique persons are being served overall, by the various programs and the overlap of the same person being served in one or more programs.
- **Extract, Transform, and Load (ETL):** DOHMPI loads data in both real-time and batch modes ensuring up-to-date identity linkage from contributing systems. DOHMPI also standardizes data across different systems making it easier to share.
- **Push and Pull Messaging Workflows:** DOHMPI supports data pull (queries) and push workflows between systems in both real-time and batch modes. This means information that is being shared can be pushed (similar to a notification) or pulled (similar to performing a search). Having the ability to both push and pull data provides the most flexibility for systems that can be customized based on a specific workflow. Workflow examples include:
    - **Data Notification:** DOHMPI provides timely program-specific notifications to destination systems. Systems only receive notifications for persons already in their system, which reduces unnecessary data transmission.. For example, if System A received death notifications from the death registry, and John Doe passes away, his death record would only be forwarded to System A if he already existed in that system.
    - **Person-Linked Data Warehousing:** DOHMPI simplifies the integration of systems sending data to a Data Warehouse by using the hub design. Identity (people) matching is done upfront by DOHMPI and pseudo-identifiers can be sent to identify a person and only include limited demographics (e.g. birth month and year, zip code). This means less identifiable information can exist in the Data Warehouse which helps address many privacy and security concerns. An advantage of a Person-linked Data Warehouse is that it does not require person matching every time a research project is performed, because the person matching has already been completed. Also the person linkage is continually updated as data is sent to the Data Warehouse.

Application Programming Interface (API): In addition to integrating with database connections or manual extracts, DOHMPI provides a secure API for connected systems. Instead of sending demographic data for matching to every applicable system, the user only needs to send their demographic data to a single system (the hub, DOHMPI API) for matching purposes. Also, instead of creating many point-to-point connections (see Graph 2 - Point-to-Point) for data sharing, systems only need to maintain a single connection to the hub (DOHMPI API) as the hub is already connected to all systems.

DOHMPI Connected Programs *(as of November 2021)*

The following programs are integrated with DOHMPI. A separate system that currently leverages the DOHMPI infrastructure and functionality is a UDOH person-linked Data Warehouse called Early Childhood Integrated Data System (ECIDS). Table 1 provides a summary of programs who are integrated with DOHMPI and who also send data to ECIDS:

Table 1 - Programs connected to DOHMPI and ECIDS status (Production)

	DOHMPI Code	Data Source Program(s)	ECIDS Integration
1	ASQ	Ages and Stages Questionnaire (ASQ) Screening	Yes
2	BTOTS	Early Intervention Part C (Baby Watch)	Yes
3	CCS	DWS- Child Care Subsidy	Yes
4	CSD	DOPL-Controlled Substance Database	No
5	EDEN	Utah Death Registry	Yes
6	HS_Centro	Head Start-Centro de la Familia	Yes
7	HFD	Healthcare Facilities Database (inpatient and outpatient encounters)	No
8	HiTrack	Newborn Hearing Screening	No
9	Medicaid	Medicaid	No
10	OHV Legacy	Office of Home Visiting (legacy system)	Yes
11	OHVP	Office of Home Visiting (current system)	Yes
12	Professional Licensing	DOPL-Professional Licensing	No
13	UCR	Utah Cancer Registry	No
14	Uintah	Utah Birth Registry	Yes
15	USIIS	Utah State Immunization Information System	No
16	WIC	Women, Infant, Children	Yes

The following programs are being actively integrated with DOHMPI:

Table 2 - Programs in process of being connected to DOHMPI and ECIDS status (In Progress)

DOHMPI Code	Data Source Program(s)	Planned ECIDS Integration
APCD	**All Payer Claims Database (Health Insurance Payers)**	**No**
TBIP	**Traumatic Brain Injury**	**No**
HS_DDI	**Head Start - DDI Vantage**	**Yes**

EFFICIENCIES, IMPACTS, AND PROCESS CHANGES

The major efficiencies, impacts, and process changes of DOHMPI are having a single information sharing hub infrastructure that can be used for both operational and research purposes. By using DOHMPI, participating systems avoid duplicating the services listed below that DOHMPI provides, which allows us to complete tasks more timely and to complete tasks that we may not have been able to complete without this resource.

1. Person matching
2. Extract/transform/data loading (ETL)
3. Message routing
4. Reduce the number of point-to-point connections to the same system.
5. Keeping system specific data (e.g. actual immunizations, birth details) in the source system, instead of multiple subset copies of the data being in multiple system(s).
6. Person-Linked Data Warehouse

The Department of Health runs quality assurance (QA) monthly on the DOHMPI data to validate that persons are linked accurately. The QA process tries to find persons that should have been linked but were not by DOHMPI (false negative) and persons who were linked that should not have been (false positives). The QA process has found that the percentage of true positives is consistently above 99.9% and the percentage of true negatives is above 90%. The Department provides feedback from the QA process to the DOHMPI vendor so the algorithms can be continuously improved.

Identity Matching

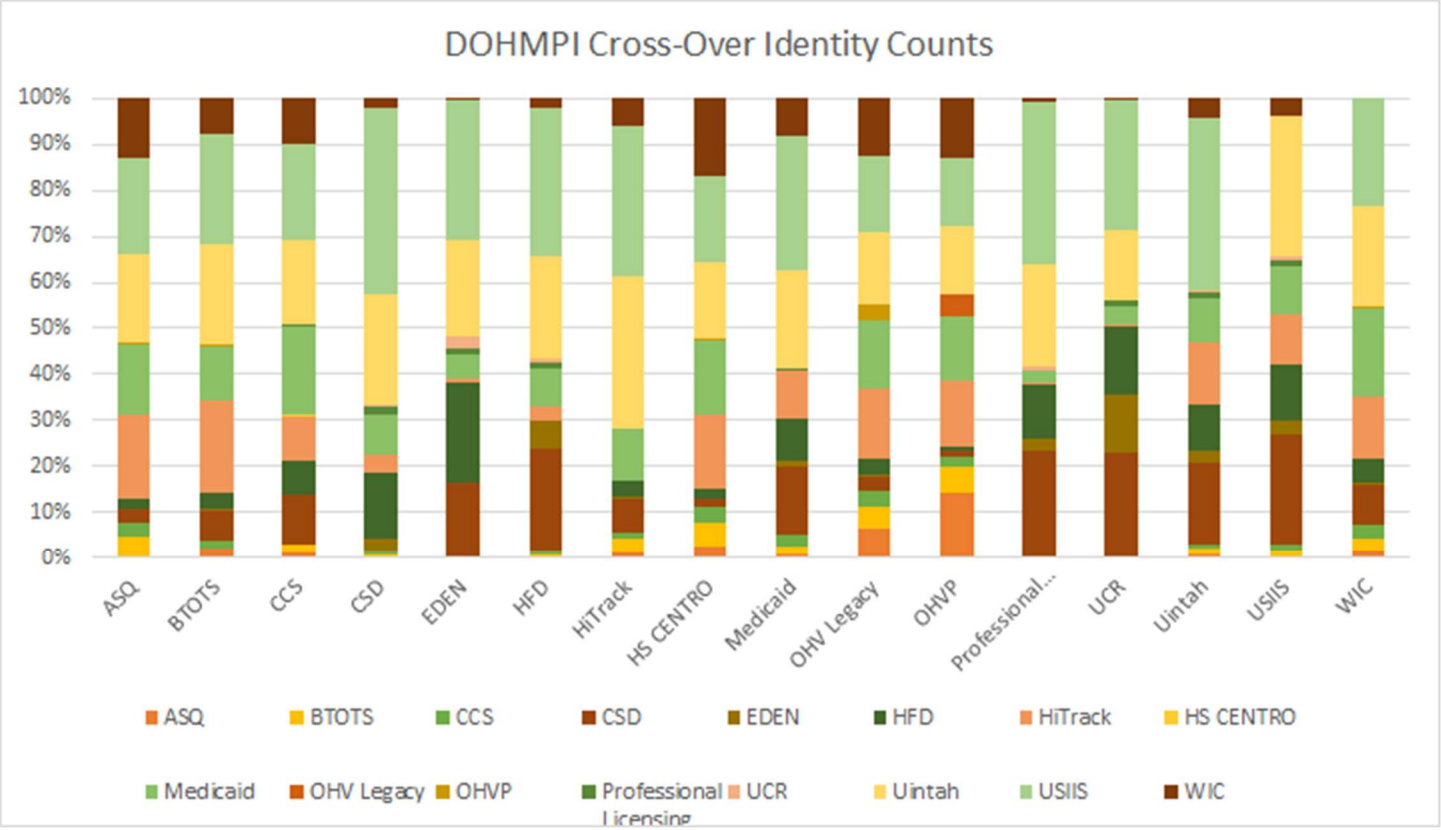
DOHMPI contains just over 10 million unique people across all programs. The number of people if counted within each program system silo would be just over 21 million people. Many people exist in two or more programs and DOHMPI can identify and report on these relationships (see **Table 3**). This table communicates the unique person count relationship between programs. **Example:** The Programs Medicaid and BTOTS contain 52,198 common unique people between them.

Table 3 – Same Person Counts Across Programs

Systems	ASQ	BTOTS	CCS	CSD	EDEN	HFD	HiTrack	HS CENTRO	Medicaid	OHV Legacy	OHVP	Professional Licensing	UCR	Uintah	USIIS	WIC
ASQ	-	7,914	4,755	4,954	298	3,241	30,820	146	25,112	516	722	2	14	31,765	34,664	21,625
BTOTS	7,914	-	7,968	29,822	1,321	15,454	90,460	299	52,198	374	312	57	101	97,931	107,215	33,565
CCS	4,755	7,968	-	47,794	1,023	31,859	44,225	189	86,612	272	103	1,175	192	80,945	94,072	42,828
CSD	4,954	29,822	47,794	-	139,134	753,744	213,413	92	462,343	251	65	93,605	38,544	1,272,351	2,145,959	116,780
EDEN	298	1,321	1,023	139,134	-	192,327	9,066	1	46,207	14	7	10,499	21,205	184,934	266,232	2,559
HFD	3,241	15,454	31,859	753,744	192,327	-	108,580	119	278,097	289	49	47,393	25,119	750,141	1,072,191	71,326
HiTrack	30,820	90,460	44,225	213,413	9,066	108,580	-	923	333,215	1,231	729	1,425	445	976,000	967,404	178,392
HS CENTRO	146	299	189	92	1	119	923	-	929	7	10	-	1	945	1,084	948
Medicaid	25,112	52,198	86,612	462,343	46,207	278,097	333,215	929	-	1,173	713	11,620	6,753	673,582	911,091	258,363
OHV Legacy	516	374	272	251	14	289	1,231	7	1,173	-	259	-	1	1,279	1,314	989
OHVP	722	312	103	65	7	49	729	10	713	259	-	-	-	744	773	668
Professional Licensing	2	57	1,175	93,605	10,499	47,393	1,425	-	11,620	-	-	-	2,540	91,010	143,007	3,161
UCR	14	101	192	38,544	21,205	25,119	445	1	6,753	1	-	2,540	-	25,744	47,867	435
Uintah	31,765	97,931	80,945	1,272,351	184,934	750,141	976,000	945	673,582	1,279	744	91,010	25,744	-	2,731,338	288,941
USIIS	34,664	107,215	94,072	2,145,959	266,232	1,072,191	967,404	1,084	911,091	1,314	773	143,007	47,867	2,731,338	-	313,179
WIC	21,625	33,565	42,828	116,780	2,559	71,326	178,392	948	258,363	989	668	3,161	435	288,941	313,179	-

Figure 3 contains a visual graph of Table 3 data.

Figure 3 – Same Person Counts Across Programs Graph



Use Case Examples

Early Childhood Integrated Data System (ECIDS) - An early childhood integrated data system (ECIDS) collects, integrates, maintains, stores, and reports information from early childhood programs within a state that serve children and families from birth to age 8. Utah is one of many states with an ECIDS system[1]. The purpose of an ECIDS is to provide integrated data across many programs that inform decisions about early childhood policies, services, and education.

ECIDS is a person-linked Data Warehouse that contains 'limited' demographic and Program specific data for children. For example, names and street addresses are not kept in the Warehouse. Person matching is performed by DOHMPI before data is sent to ECIDS. ECIDS maintains their own website (https://ecids.utah.gov/) for users to interact with the data. ECIDS leverages DOHMPI infrastructure for the following services:

- Real-time Identity Resolution and Record Linkage
- Extract, Transform, and Load (ETL)
- Push and Pull Messaging and Workflows
- Identity (person) Linked Data Warehousing

ECIDS would cease to exist without DOHMPI infrastructure and services. Replacing the functionality that DOHMPI provides for ECIDS would require significant one-time and on-going capital requirements and resources that ECIDS does not have.

Population Health Workforce Initiative Project - A use case being developed is leveraging DOHMPI to match persons across the following programs' data and over time: Healthcare Facilities Database (HFD) which contains information on hospital admissions, Eden which is Utah's Death Registry, and the All Payer Claims Database (APCD Processed) which houses information on insurance claims. The goal is to research, measure, and better understand mental illness in Utah and the associated costs.

Medicaid Eligibility Data Linkages - Medicaid eligibility data has been integrated into DOHMPI, in the form of their All Payer Claims Database files (Medicaid APCD). Medicaid and UDOH planned to link this data to 15 different datasets which can be used to analyze cost effectiveness, research comparative effectiveness, and measure outcomes. Of the 15 planned datasets, 10 are integrated into DOHMPI and 2 more are in the process of being integrated.

Perinatal/Utah Birth Defects Network Analysis - Another use case that is being explored is leveraging DOHMPI to match children for research studies over time across all of the ECIDS Programs and the Utah Birth Defects Network. Currently, the Perinatal and the Utah Birth Defects Network programs can do

[1] https://slds.ed.gov/#program/ecids-toolkit
https://ecdwcommunity.org/

analysis on their own data and a subset of data from the Utah Birth Registry. The goal is to identify additional Utah Birth Registry risk factors to send to ECIDS so studies can be performed that analyze birth risk factors and identify patterns as it relates to screening result outcomes. In addition, other programs may be able to identify patterns based on birth risk factors as it relates to their data.

Data Notification(s) - DOHMPI provides new data notification functionality that can reduce manual work and be extended quickly and easily to more systems. Examples include:

- **Death Notifications** - Programs who are authorized to receive the data under law are able to be automatically notified when a death occurs with a person in their system. This helps replace manual processes in looking up and/or monitoring **ALL** Utah deaths.
- **Duplicate Identifiers for the same person within a single Program Notification** - Programs can be automatically notified when duplicate identifier(s) for the same person exists in their system.
- **Same person exists across two or more Program(s) Notification** - Programs are able to be notified when the same person exists in two or more Programs systems.

Controlled Substances Database (CSD) Unique Persons - CSD assigns unique person identifiers whenever there are distinct values on specific data elements, because names may be typed differently by different providers they often create multiple identifiers for the same person. By linking multiple CSD person identifiers into a single identity, related controlled substances/prescriptions can be accurately associated to a single person without manual intervention from a medical provider, which is required today. Using DOHMPI's matcher to identify duplicate records, DOHMPI has reduced the number of unique CSD people by 53%. DOPL plans to utilize DOHMPI identity resolution in the near future for their dashboard reporting. Further analysis is needed in how to incorporate DOHMPI identity resolution for their operational system that providers use.

All Possible Traumatic Brain Injury (TBI) Unique Persons - A TBI registry is being created for persons with TBI injuries so outreach can be done to help assist individuals to utilize TBI related funding and resources. The HFD database contains inpatient and outpatient admissions reported by different organizations, multiple hospitals, ambulatory and emergency department environments. A subset of these records may involve visits associated with TBI related admissions. The people involved in these visits may not have enrolled in the TBI Registry.

An integration is being explored to leverage DOHMPI to match identities between the Healthcare Facilities Database (HFD) database and the TBI registry as allowed under R486-703-4(4)(e) if the health care facility requests reporting through this mechanism so that they don't have report separately. DOHMPI will link the identities across these Programs and identify individuals who had TBI related visits that are not in the self-enrolled TBI registry so outreach can be done.

COST IMPACTS

CURRENT FUNDING SOURCES FOR DOHMPI

Current funding sources of the DOHMPI system and activities include State general funds, Centers for Medicare and Medicaid Services (CMS) funds obtained through an Operational Advanced Planning Document, the Overdose to Action federal grant through the Centers for Disease Control and Prevention, and the State Traumatic Brain Injury Fund. In State fiscal year 2021, the DOHMPI system was funded 68% ($357,632) by CMS, 21% ($109,400) by State general funds, 9% ($46,008) by the Overdose to Action grant and 2% ($11,600) by the TBI fund. The general fund and CMS funds are ongoing. The Overdose to Action grant ends on 4/30/2022 and the TBI funding ends on 8/31/2022.

CONCLUSION/RECOMMENDATIONS

Prior to DOHMPI, when data needed to be linked across multiple programs, basic algorithms and manual processing was often used. This can be time consuming and needs to be repeated each time a use case arises. DOHMPI's more sophisticated person matcher is up-to-date and can be used at any time. In addition, significantly less manual intervention is required, making it a more scalable solution. Having more automated processes to look at unique persons across multiple systems provides Utah the opportunity to view and understand the complete picture of programs being used by individuals.

An example of this is helping to better understand how the same person is enrolled in programs, their level of involvement in the program(s), how they transition from one program to another program and how many programs the same individual is involved in at the same time. The Early Childhood Integrated Data System (ECIDS) Data Warehouse is doing this and can be a model for a new DHHS person-linked data warehouse that consists of already matched persons. The more data you have in a single person-linked data warehouse, the more comprehensive analysis you can perform. Baseline measurements and changes over time can be measured with real data.

DOHMPI is being used by UDOH and external Data Partners (e.g. DWS, DOPL and private sector organizations). DOHMPI can also be leveraged by the new DHHS for their planned data warehouse.

The DOHMPI information hub infrastructure and services is an on-going project that has already started to see significant efficiencies and introduce new functionality that is not possible within a single program specific data silo, these include but are not limited to:

- The same data sharing infrastructure (DOHMPI) can be leveraged for research and operational purposes for all programs.
- Simplify and significantly reduce the integration barriers of data sharing across multiple program system data silo(s).
- Reduce the administrative costs and complexity associated with point-to-point data connections and integrations.
- Not having to invest or have the resources to maintain duplicate and/or program system specific hub sharing infrastructure information silos.
- Can query a single system (DOHMPI) to locate which program systems contain data regarding an individual.
- The infrastructure to create a department-wide Data Warehouse where people matching across all programs is done upfront and one time versus every time a different use case is needed (e.g. similar to what ECIDS is already doing):
    - Studies could be performed across some or all programs. To look at unique persons across program specific silo(s) and how that person's data may be spread across multiple program data silo(s) and begin to understand how and why they may interrelate.
    - Helping to better understand how the same person is enrolled in programs, their level of involvement in the program(s), how they transition from one program to another program and how many programs the same individual is involved in multiple programs at the same time
    - Identifying individuals who could benefit from various programs at an earlier age based on their risk factors, who are not currently enrolled in programs that may assist in overcoming and/or dealing better with known risk factors. Helping the individual, the family, the community and save the public/taxpayers money.
    - UDOH (and the new DHHS) may be able to to better initiate the discussion (based on data) and could have more direct influence in and leading the collaboration with the community to analyze, implement changes and measure the impact(s) of these changes placing UDOH (and the new DHHS) in a more proactive role (e.g. being able to do root cause analysis, make changes, measure impact and have continuous quality improvement) that can affect outcomes.

Recommendation:

The recommendation is to continue to fund DOHMPI and make the appropriate investment to further develop the DOHMPI Data Sharing Hub so the efficiencies and new functionality can be further realized. This includes but is not limited to:

- Integrate more programs with DOHMPI.
- Further develop and have more programs leverage the DOHMPI API for data sharing.
- Explore using DOHMPI infrastructure in creating a department-wide Person-Linked Data Warehouse where people matching across all programs is done upfront before data is loaded versus every time a different use case is needed.

All of the investment of time, resources and funding made to DOHMPI could also be utilized by the new DHHS. DHHS can integrate more programs with DOHMPI, leverage the DOHMPI API for data sharing, and create a more robust and complete department-wide Person-Linked Data Warehouse where people matching across all programs is done upfront before data is loaded.